IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICANT	: Jackowski *et al.*
INVENTION	: **Glycoprotein and Apolipoprotein Biopolymer Markers Indicative of Alzheimer's Disease**
SERIAL NUMBER	: 09/993,344
FILING DATE	: November 23, 2001
EXAMINER	: Chernyshev, Olga N
GROUP ART UNIT	: 1649
OUR FILE NO.	: 2132.096

CERTIFICATE UNDER 37 CFR 1.8(a)
I hereby certify that this correspondence is being deposited with the U.S. Postal Service as First Class mail in an envelope addressed to Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 on 10-11-2005

Mail Stop: Non-Fee Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 CFR § 1.132

I, Ferris H. Lander, do hereby declare as follows:

1. I am a registered Patent Agent and am authorized to represent the inventor's and assignee in the application entitled **"Glycoprotein and Apolipoprotein Biopolymer Markers Indicative of Alzheimer's Disease"**, having U.S. Application Serial No. 09/993,344, filed November 23, 2001.

2. In the Office Action mailed on July 28, 2005, claim 1 (as presented on April 4, 2005) was rejected under 35 USC 101 because the claimed invention allegedly is not supported by either a

specific, substantial, credible or asserted utility or a well-established utility. Claim 1 was also rejected under 35 U.S.C. 112, first paragraph because the claimed invention allegedly contains subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention.

3. The attached figure was produced by scanning the original photograph of the gel. No new matter has been added; this figure is simply a clearer copy of Figure 1 as originally filed and is provided for clarification of the presence and/or absence of the bands. The figure entitled "DEAE 3 (Elution) AD vs. Age Matched AD(Control)" represents Figure 1.

The gel shown in the attached figure does not represent new experimentation; the figure shows a clearer image of the original gel made at the time that the experiments described in the instant specification were first carried out.

The undersigned declares that all statements made herein of his own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the Application or any patent issuing thereon.

12/11/2005
Date

Ferris H. Lander
Ferris H. Lander
Reg. No. 43,377

\\Ns2\SERVER\CLIENT FILES\2100-2199\2132 -Syn-X\2132_000096 - Glyocoprotein and Apolipoprotein\Amendments\2132_096.132.wpd

9132.096
Examiner Copy
reference #1

1: J Neural Transm. 1996;103(4):433-46. Related Articles, Links

Elevated 5-S-cysteinyldopamine/homovanillic acid ratio and reduced homovanillic acid in cerebrospinal fluid: possible markers for and potential insights into the pathoetiology of Parkinson's disease.

Cheng FC, Kuo JS, Chia LG, Dryhurst G.

Department of Medical Research and Geriatrics Medical Center, Taichung, Taiwan, Republic of China.

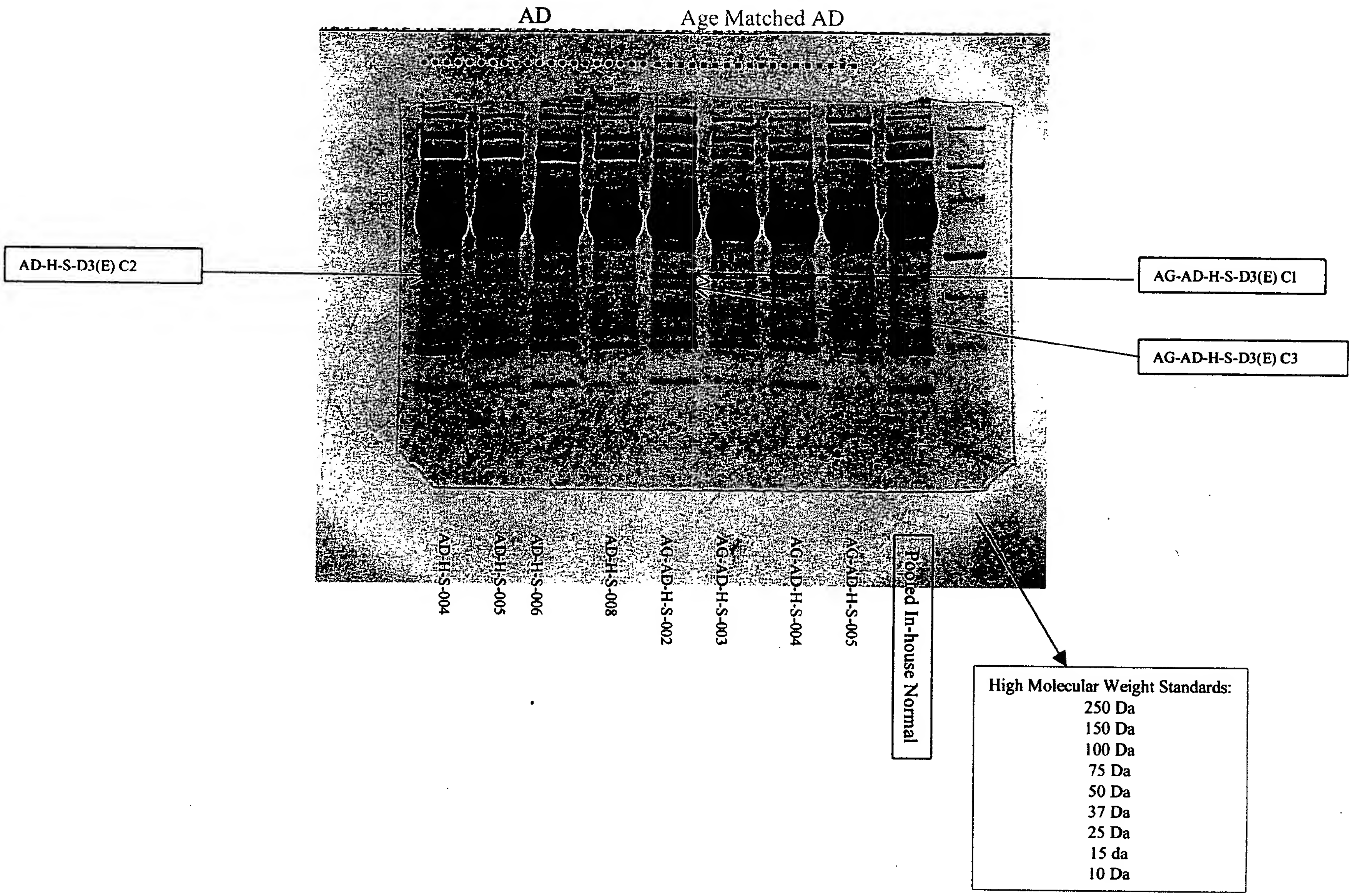
High-performance liquid chromatography with electrochemical detection has been employed to analyze ultrafiltrates of cerebrospinal fluid of Parkinson's Disease (PD) patients and age-matched controls for the dopamine (DA) metabolites homovanillic acid (HVA) and 5-S-cysteinyldopamine (5-S-CyS-DA). The mean level of HVA in the CSF of PD patients, measured 5 days after withdrawal from L-DOPA therapy, was significantly lower than that measured in controls. By contrast, mean levels of 5-S-CyS-DA were not significantly different in the CSF of PD patients taking L-DOPA (PD-LT patients) the same patients 5 days after discontinuing this drug (PD-LW patients) or controls. However, the mean 5-S-CyS-DA/HVA concentration ratio was significantly ($p < 0.05$) higher in the CSF of PD-LW patients compared to controls. Although the PD patient population employed in this study had been diagnosed with the disease several years previously and had been treated with L-DOPA for prolonged periods of time the results of this study suggest that low CSF levels of HVA and a high 5-S-CyS-DA/HVA ratio together might represent useful markers for early diagnosis of PD. The high 5-S-CyS-DA/HVA ratio observed in the CSF of PD-LW patients also provides support for the hypothesis that the translocation of glutathione or L-cysteine into neuromelanin-pigmented dopaminergic cell bodies in the substantia nigra might represent an early event in the pathogenesis of PD.

PMID: 9617787 [PubMed - indexed for MEDLINE]

Display Abstract Show 20 Sort by Send to

Figure 1
2132.096

DEAE 3(Elution) AD vs. Age Matched AD (Control)



2132.096
Examiner copy
reference #2

link[1] ngk)
n.

One of the rings or loops forming a chain.

A unit in a connected series of units: *links of sausage; one link in a molecular chain.*

A unit in a transportation or communications system.

A connecting element; a tie or bond: *grandparents, our link with the past.*

An association; a relationship: *The Alumnae Association is my link to the school's present administration.*

A causal, parallel, or reciprocal relationship; a correlation: *Researchers have detected a link between smoking and heart disease.*

A cuff link.

Abbr. **li** A unit of length used in surveying, equal to 0.01 chain, 7.92 inches, or about 20.12 centimeters.

A rod or lever transmitting motion in a machine.

Computer Science. A segment of text or a graphical item that serves as a cross-reference between parts of a hypertext document or between files or hypertext documents. Also called **hotlink, hyperlink**.

as accessed from the internet
on October 7, 2005

Mass spectrometry and proteomics

2132.096
Examiner copy-
reference #. 3-

SCOTT D. PATTERSON
Amgen Inc., Thousand Oaks, California 91320-1789

THE 15TH ASILOMAR CONFERENCE on Mass Spectrometry this October was devoted to the role of mass spectrometry (MS) in proteomics. The Asilomar Conference site is in a picturesque national park in Pacific Grove, CA, overlooking the Pacific Ocean. The conference aims to bring together scientists from a cross section of disciplines that are applying MS to an emerging field. This year, that emerging field is proteomics. The term "proteome" was coined by Wilkins et al. (17) in the mid-1990s to describe the protein complement of the genome. The term was first used to describe the 20-yr-old field of two-dimensional gel electrophoresis (2-DE) and quantitative image analysis. 2-DE remains the highest resolution protein separation method available, but the ability to identify the observed proteins was always an extremely difficult problem. MS has been integral to solving that problem. Although improvements in 2-D gel technology had been realized since its introduction, three enabling technological advances have provided the basis for the foundation of the field of proteomics. The first advance was the introduction of large-scale nucleotide sequencing of both expressed sequence tags (ESTs) and, more recently, genomic DNA. The second was the development of mass spectrometers able to ionize and mass-analyze biological molecules and, more recently, the wide-spread introduction of mass spectrometers capable of data-dependent ion selection for fragmentation (MS/MS)(i.e., without the need for user intervention). The third was the development of computer algorithms able to match uninterpreted (or partially interpreted) MS/MS spectra with translations of the nucleotide sequence databases, thereby tying the first two technological advances together. Thus MS played a key role in the passage of 2-DE/image analysis to proteomics.

As a note to readers unfamiliar with MS, the instruments are named for their type of ionization source and mass analyzer (see also Refs. 1, 11, 12). To measure the mass of molecules, the test material must be charged (hence ionized) and desolvated (dry). The two most successful mechanisms for ionization of peptides and proteins are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). In MALDI the analyte of interest is embedded in a matrix that is dried and then volatilized in a vacuum under ultraviolet laser irradiation. This is a relatively efficient process that ablates only a small portion of the analyte with each laser shot. Typically, the mass analyzer coupled with MALDI is a time-of-flight (TOF) mass analyzer that simply measures the elapsed time from acceleration of the charged (ionized) molecules through a field-free drift region. The other common ionization source is ESI, in which the analyte is sprayed from a fine needle at high voltage toward the inlet of the mass spectrometer (which is under vacuum) at a lower voltage. The spray is typically either from a reversed-phase HPLC (RP-HPLC) column or a nanospray device (19) that is similar to a microinjection needle. During this process, the droplets containing analyte are dried and gain charge (ionize). The ions formed during this process are directed into the mass analyzer, which could be either a triple-quadrupole, an ion trap, a Fourier-transform ion cyclotron resonance (FT-ICR), or a hybrid quadrupole TOF (Qq-TOF) type.

This Asilomar meeting provided one of the largest academic forums in the United States for the presentation and discussion of MS as it is applied to proteomics. As is obvious from the introduction, the initial role MS played was as a protein identification and characterization methodology. However, the role of MS is expanding in this field. Although a series of talks focused on the use of different kinds of MS to identify gel-separated proteins and the various automation technologies applied to perform this in high throughput, several talks also presented alternate approaches. These approaches utilized direct analysis of digested protein mixtures for either identification of the components or quantitative analysis of two different samples mixed together. Specific biological applications were also presented. As described above, a critical component of any MS approach as applied to proteomics is the computational analysis. This report will be divided to focus on these six aspects of MS in proteomics. Where references are known for some of the material presented, they are cited. The program was, however, not entirely limited to MS in proteomics. Prior to the six sections covering the conference core, the first section of this report covers those presentations that were aimed at providing an insight into broader biological and drug discovery processes.

Proteomics in biology and drug discovery. The opening lecture, given by Lee Hood (Univ. of Washington), provided an excellent overview of Genomics, Proteomics, and Systems Biology. Hood described the genome project efforts that provide four types of maps: genetic, physical, gene, and sequence. For the human genome,

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

it is anticipated that 90–95% of all genes will be sequenced sometime next year. This is the first step toward what Hood described as the "Periodic Table of Life." The different approaches to genomic sequencing and microarray technologies that are able to interrogate the mRNA levels of thousands of genes at a time were described. Hood described proteomics in broad terms as the study of multiplicity of proteins. The information obtained from the various hierarchical levels of biological information (gene, protein, pathways, interconnecting pathways) must be integrated for us to be able to provide a more complete biological picture. For both microarray and proteomics, samples representing the disease process must be obtained. This means that pure cell populations must be microscopically captured from tissues and/or sorted prior to analysis. Therefore, analyses at the mRNA and protein level must be conducted at very low levels and substantial engineering opportunities exist in the biological field to provide the necessary solutions. However, generation of the data is only the first hurdle, as the analysis of the data from a systems perspective then must be undertaken. Hood presented systems biology as the challenge for the 21st century and provided a series of examples of large-scale approaches to biology, from genome sequencing of unicellular organisms, to the sequencing of the T-cell receptor locus, to cancer biology, all of which benefit from such approaches.

Three other presentations were included in the program, to provide a broader background to the utilization of proteomics in drug discovery. Doug Buckley (Exelixis) described the generic view of the drug discovery pipeline, the various "choke" points in the process, and where proteomics could play a role. Of note was the discussion of the changing patent protection landscape, during which Buckley said that full-length cDNA patents were being issued despite the existence of EST patents on portions of these genes. Buckley also predicted that functional data is expected to be required for patents beyond the inferences gained from bioinformatics. The choke points he referred to were target validation, assay development, mechanistic biology, and toxicology. Exelixis is using model organisms (*Caenorhabditis elegans*, *Drosophila*, mouse, and zebrafish) to screen for genes that disrupt/modulate pathways common between man and these organisms. Roles for proteomics included follow-up on targets (direct analysis of protein differences, proteins associated with gene products of interest), assay development [e.g., validation of hits in high-throughput screening (HTS)], and mechanistic biology (e.g. comprehensive analysis of a knockout phenotype). Most importantly, Buckley presented the bottom line that all new technologies must demonstrate their worth by concrete changes in the drug development pipeline (i.e., greater efficiency, better decisions). He predicted that proteomics could provide these benefits at the multiple restriction points referred to above.

Pharmacoproteomics, using 2-DE to profile mechanisms of drug efficacy and toxicity, was presented by Tina Gatlin (Biosource/Large Scale Biology Corporation). The synergy between mRNA expression profiling (for low-abundance gene products) and protein expression profiling (for posttranslational modifications and subcellular localization) was presented. An exception to this is the search for surrogate markers, where secreted proteins were normally the choice and in which there is no identifiable mRNA source to mirror serum or urine protein expression. The aim of their Molecular Effects Database of 2-DE patterns, obtained from livers of drug-treated rats, is to establish links between expression patterns and toxic endpoints to reveal markers for efficacy and prediction of side effects which can be used for lead selection. In disease models, the hypothesis is that the altered expression pattern could be reversed by treatment with a drug.

The closing presentation of the meeting, given by Jeff Seilhamer (Incyte), presented analyses of the precursor to proteins, mRNA. The staff at Incyte have generated very large EST libraries and from these have estimated the number of genes in the human genome to be 129,769 (based on CpG island estimates, 142,634). They are now sequencing the human genome at a rate of about 1 million reads a month on the Megabace platform with 9 sequencing runs/day. Assembly of the data is being accomplished using Linux on 1,500 CPUs (160 computers) with 75 terabytes of storage capacity. Single-nucleotide polymorphisms (SNPs) are being calculated from their sizable EST collection, and mRNA expression profiling is being achieved using their GEM microarray platform. These data are being integrated with 2-DE proteomics data being generated by their partner Oxford GlycoSciences. This integration of the technologies of genomics and proteomics forms the basis of their drug discovery approach for profiling differences between normal and diseased tissue.

Computational aspects of proteomics. Determining the masses of peptides (MS spectra) derived from enzymatic digestion of gel-separated proteins is often the first step in a mass spectrometric-based protein identification strategy. Peptide-mass mapping is the most commonly employed mass spectrometric approach for protein identification from organisms whose genome is completely sequenced (or at least for which the more abundantly expressed genes have been sequenced). The basis of the method is the matching of experimentally determined peptide masses with peptide masses calculated for each entry in a sequence database (using the specificity of the enzyme used to generate the experimental data). How well the experimentally determined masses match with the calculated masses forms the basis of the approach. Ron Beavis (Proteometrics) described how to obtain high-quality data, which even if less, are better than more low-quality data. The use of specific matrices as well as the use of standards with respect to obtaining appropriate data sets for peptide-mass mapping was addressed. Later in the day David Fenyo (Proteometrics) described how to utilize this data in a three-step process as is performed in their WWW-available program, Profound, which uses a Bayesian algorithm (**http://www.proteometrics.com**). The process is as follows:

1) assignment of monoisotopic masses to the raw data, *2)* peptide-mass search, and *3)* significance testing of the result (4). The last step was presented as the most critical because it is from this that the confidence of the match is derived. This is achieved through calculation of a score frequency function for false positives. This was derived from statistical analysis of the database being searched using random selections of peptide masses from different proteins that are then grouped as synthetic proteins and used in a peptide-mass search of the database in question. This is repeated for a variety of random selections to come up with robust statistics for false positives.

The next level of protein identification is the generation of fragment ion spectra from peptides isolated in the gas phase of the mass spectrometer (MS/MS spectra). Matching of fragment ion spectra follows the same principle as for peptide-mass mapping. Experimentally calculated masses of fragment ions (together with the intact mass of the peptide, and often the specificity of the enzyme used to generate the peptide) are matched with those calculated for isobaric peptides (i.e., same mass as experimentally determined) from entries in sequence databases. Arthur Moseley (Glaxo Wellcome) described how nanoscale capillary LC-MS/MS (where peptides are separated chromatographically before MS/MS) had been automated for identification of gel-separated proteins. The throughput of this 75 μm ID capillary system connected to a Qq-TOF mass spectrometer was 20 samples per day at levels to 30 fmol (loaded on gel) for BSA. Moseley continues to develop ultra-HPLC (in some cases combined with variable flow systems) that improve both the speed and resolution of separation. In a Glaxo Organelle Proteomics program, various approaches to protein identification were examined. A comparison of the total number of proteins identified following in situ enzymatic digestion of proteins separated by either high-resolution 2-DE or one-dimensional (usually SDS-PAGE) gel electrophoresis (1-DE) was presented. Only one or a limited number of proteins are present in each of the 2-DE spots, whereas many proteins were present in the 1-DE bands of the enriched Golgi complex. In fact, more proteins were identified from the 1-DE bands than from the 2-DE spots (see below, *Analysis of complex protein mixtures without gel electrophoresis*).

MS/MS spectra derived from tryptic digestions conducted in the presence of equal quantities of $H_2{}^{16}O$ and $H_2{}^{18}O$, when combined with subtractive analysis of the two types of spectra, allows de novo sequencing as described by Matthias Wilm (EMBL) (18). By utilizing a Qq-TOF mass spectrometer, peptides containing both COOH-terminally incorporated stable isotopes and just the isoform containing the ^{18}O could be selected for fragmentation from the mixture. Subtraction of the ^{18}O spectrum from the ^{16}O:^{18}O spectrum reveals only ^{16}O *y*-series ions. Often, a complete ion series is obtained. The method has proved feasible in their hands when 1 pmol of protein is present in the gel (1/4 of this amount can be successfully analyzed with standard digest conditions).

Automated identification of gel-separated proteins by mass spectrometry. Following quantitative analysis of 2-DE patterns, the next step is the identification of all protein spots that display differential expression. Andrew Gooley (Proteome Systems) described the approaches they are employing for quantitative analysis using 2-DE. This included the following: sample preparation (sequential detergent extraction with aminosulfobetaine-14), narrow-range immobilized pH gradient (IPG) with mini-gels for the 2nd dimension, through to the robotic system that they have codeveloped for spot excision, liquid handling (peptide extraction and reverse-phase bead cleanup and storage) and peptide-mass fingerprinting by MALDI-MS. Apart from the throughput of the robotic system, diminished contamination from keratin and more reproducible spotting of samples for MALDI-MS is a highly desirable feature of automation. Hans-Werner Lahm (Hoffmann-La Roche) described the high-throughput system they use for automated spot excision from 2-DE, digestion (with low-salt buffer to eliminate the need for cleanup), and spotting for automated MALDI-MS. Lahm also described the computational aspects of operating such a system in high-throughput mode for long periods of time, including automated database search routines for users distributed throughout the world at other Roche sites. They are investigating the use of stable isotope labeling (^{14}N/^{15}N) followed by mixing of each sample prior to 2-DE for direct quantitation of relative expression differences from the MALDI-MS spectra of individual protein spots. The system averages 1,000 spots to spectra per day (including downtime).

David Arnott (Genentech) described automation of in-gel digestions following analysis of differentially regulated proteins from 2-DE. Arnott described the trapping cartridge approach that was required to analyze extracted peptides from the DigestPro robot (currently 30 sample spots, but upgradeable to 96) by microcapillary LC-MS/MS. They aimed to automate as much of the sample processing as possible with automated liquid handling from the digestion robot to the data-dependent LC-MS/MS (capable of handling 40 samples per day) using an ion-trap mass spectrometer followed by auto-database searching using Sequest (3). The system is capable of analysis of subpicomolar quantities of protein from silver-stained gels.

Advances in separations and mass spectrometers. Accurate mass analysis of intact proteins using an 11.5-T FT-ICR coupled with a capillary electrophoresis (CE) instrument was demonstrated by Richard Smith (Pacific Northwest National Laboratory) as a means of proteome analysis. Through the use of stable isotope labeling of one sample and running that sample with an unlabeled sample provides the possibility to measure protein expression ratios. To identify the proteins that display different ratios, dissociation in the FT-ICR-MS to yield mass tags is possible. Having intact mass information as well as identification allows post-translational modifications to begin to be investigated. The mass accuracy obtainable by this FT-ICR-MS was said to be <0.75 ppm which allows the generation of

accurate mass tags for tryptic peptides. In many cases this may be sufficient for protein identification (at least for an organism like *C. elegans*). In some cases, MS/MS may be required, but once performed it would not have to be repeated. Another possibility is the identification of cysteine-containing peptides at a mass accuracy of 1 ppm, which was said to be sufficient for identification. Another possibility being explored with this instrument is multiplexed MS/MS, where up to 7 ions could be isolated at once and the MS/MS spectrum could be deconvoluted for each selected ion (requires accuracy of <10 ppm). This will be tried with online separations in the near future.

Marvin Vestal (PE BioSystems) described his continuing efforts in MALDI-MS instrument design. The attributes he is aiming for include sensitivity, specificity (resolution, mass accuracy, selective ionization), speed, accuracy, dynamic range, and mass range. The sensitivity will always be limited by chemical noise, but the aim is to reduce the limitations of ionization and data handling. Vestal would like to achieve a sensitivity of 1 fmol with a data acquisition rate of 1 spectrum/s. The instrument he is designing to achieve these aims is a MALDI-TOF/TOF-MS. This system has an ion gate (with 500 resolution and no loss of sensitivity) after the collision cell so that metastable ions created after reacceleration are removed. Although this system is in the early stages of development, the data shown demonstrate that this instrument is meeting most of the stated objectives.

The hybrid quadrupole TOF (Qq-TOF) mass spectrometer developed a few years ago (9), which has now been commercialized, utilizes an ESI for ionization (10). Both Ken Standing (Univ. of Manitoba) and Brian Chait (Rockefeller Univ.) described the use of a MALDI ion source for introduction of ions into a modified commercial Qq-TOF, thus taking advantage of both the high-efficiency ion production of the MALDI and the ion isolation/fragmentation of the quadrupole system with a TOF mass analyzer. Standing presented data showing sensitivity of purified standards (e.g., Substance P) in the 70 amol range (1-min acquisition) for MS and 7 fmol for MS/MS with 10,000 resolution. This instrument offers similar advantages to the TOF/TOF described above.

Online MS analysis of capillary electrophoretic or chromatographic separations of peptide (or proteins) is most often achieved using ESI-MS. Barry Karger (Barnett Institute) described how very small quantities of peptides/proteins could be separated and analyzed using vacuum deposition onto Mylar audio tape for subsequent coupled MALDI-MS analysis. The approach had so far been multiplexed with the effluent of 12 capillaries being deposited under vacuum onto the tape. The approach is designed for high-throughput separations and mass analysis.

Proteomic analyses often employ 2-DE, but David Lubman (Univ. of Michigan) described a liquid-phase 2-D separation of proteins utilizing a novel MS. The requirements of his mass spectrometer were high sensitivity, low duty cycle, and fast response. He designed and built an ion trap to capture ions from the CE coupled to TOF-MS. The 2-D liquid-phase separation consisted of nonporous silica bead RP-HPLC (which provided good resolution <50 kDa) that was conducted at high pH followed by CE and MS. Whole cell lysates were analyzed with this system, and some of the data obtained were presented.

Biological applications. Brian Chait (Rockefeller Univ.) presented the culmination of an enormous amount of work at both the protein chemistry (mass spectrometry) and cell biology levels. The nuclear pore complex (NPC) in yeast is a massive structure (1,000 Å across with 8-fold symmetry) that regulates protein transport in/out of the nucleus. The first step in understanding this structure was to purify the complex and then identify every protein present. The protein fraction was separated by several different chromatographic steps followed by SDS-PAGE from which every visible band was excised and analyzed by MALDI-IT-MS. This was an especially daunting task as the NPC when isolated contains a snapshot of the proteins transiting the NPC at that point in time. Hence, of the 174 proteins identified, 29 were nucleoporins and only 14 were shown to be present in the NPC. These 14 proteins were characterized as being present in the NPC by a variety of techniques. Protein A (4.5 repeats of the Fc binding region) fusions with the proteins of interest were generated, and immunohistochemistry was performed on cells transfected with these constructs. Electron microscopy of hundreds of NPCs following transfection allowed stoichiometry and symmetry (nuclear/cytoplasmic/asymmetric) to be determined. Subcellular fractionation and high-pH extractions were also performed to further characterize localization biochemically. This elegant study has allowed a testable model for nuclear transport to be constructed.

Two examples of the utility of analysis of unfractionated or partially fractionated complex protein mixture digests (see next section) were presented by Scott Patterson (Amgen Inc.). As a first step in the understanding of the interchromatin granule clusters (IGC), a nuclear organelle which is a major site of mRNA splicing. Samples enriched in this structure were digested with trypsin, and the complex mixture of peptides was analyzed by data-dependent LC-MS/MS (8). Some proteins known to be present in these structures were identified together with 19 novel genes (including ESTs). Three of the genes were confirmed to be present in the IGC by immunohistochemistry of cells transfected with yellow fluorescent protein (YFP)-fusion constructs with counter staining of splicing factors. The other study presented identified 108 proteins present in a protein fraction obtained from isolated mitochondria treated with atractyloside [mimicking in vitro the permeability transition pore complex (PTPC) which occurs during apoptosis] (13).

Analysis of immunoprecipitates using a new affinity strategy was presented by Gitte Neubauer (EMBL). The new strategy is referred to as tandem affinity purification (TAP) and was developed by colleagues at EMBL (15). The system utilizes a double tag for higher

Physiological Genomics

specificity and much reduced background. The human spliceosome immunoprecipitated under normal conditions (see Ref. 5 for same approach with yeast tri-snRNP) and using the TAP method were compared, demonstrating the utility of this approach.

The common theme of all of these applications is that MS was utilized early on to provide rapid and accurate protein identifications. The genes identified could then be further analyzed to attempt to determine their function.

The use of MS to identify proteins from 2-DE gels was also described by Al Burlingame (Univ. of California, San Francisco) and Reid Townsend (Oxford GlycoSciences). Burlingame described their work to identify protein targets of acetaminophen during acute toxicity and the intricacies of such analyses (14). Townsend described an Oxford GlycoSciences and Pfizer collaboration to separate by 2-DE and identify proteins from cerebrospinal fluid (CSF) in a study aimed at identifying markers for Alzheimer's disease. CSF is a compartment isolated by the blood-brain barrier but it is not just a filtrate of blood. It is produced by the choroid plexus and has a total central nervous system volume of about 90–150 ml that is turned over a few times per day. Comparative analysis of matched plasma CSF samples (in addition the normal/diseased samples) revealed that key plasma proteins (e.g., albumin, transferrin, IgG) showed markedly different relative ratios between plasma and CSF. For effective 2-DE analysis of these samples, a selective removal of albumin, IgG, transferrin, and haptoglobin had to be developed. This was accomplished by affinity depletion. Interestingly, many features in a 2-DE separation are albumin fragments (in fact, 4% of total features). Their study included 512 samples from 228 patients and resulted in 1,131 features (spots) being annotated. Potential markers of Alzheimer's disease were said to be identified.

Separate from the MS identification issues covered in most of the meeting, Kerstin Strupat (Univ. of Muenster) presented her work on MS analysis of noncovalent complexes. Here the challenge is to transfer noncovalent interactions that occur in the condensed phase to the gas phase. ESI-MS has been shown by a number of groups to work, but MALDI-MS analysis has proved more difficult. Examples of MALDI-MS analysis of noncovalent protein:protein (streptavidin tetramer and the macrophage migration inhibitory factor related proteins MRP-8 and MRP-14) and protein:ligand (aldose reductase:NADP) interactions were presented. Interestingly, analysis of the first laser pulse during a MALDI-MS analysis often allows investigation of noncovalent interactions that are not observed during subsequent pulses (16).

Analysis of complex protein mixtures without gel electrophoresis. The first stage of many proteome projects is the identification of the components comprising the system under study. This is of course the first step in understanding any biological system. As described above, an increasing (but still limited) number of laboratories have access to robotic systems requisite for the analysis of large numbers of spots from 2-DE. However, a trend in the field is emerging toward the elimination of the high-resolution protein separation step prior to protein identification by MS. In this approach, the entire enriched protein fraction is enzymatically digested (usually with trypsin), and the resulting complex peptide mixture is subjected to data-dependent LC-MS/MS. In this approach the peptides are separated by both hydrophobicity (RP-HPLC) and charge (m/z in the mass spectrometer) prior to ion selection by the MS control software (hence, data dependent). At this meeting, presentations from five groups demonstrated the utility of the approach to identify components of complex mixtures.

Analysis of immunoprecipitated proteins or enriched protein fractions (e.g., Golgi complex) by either gel electrophoresis followed by in-gel digestion and MS or digestion of the entire protein fraction and analysis by data-dependent LC-MS/MS using a Qq-TOF was described by Jyoyti Choudhary (Glaxo Wellcome). Batched MS/MS spectra were searched using the Mascot program (**http://www.matrixscience.com**). Data presented showed that if the immunoprecipitate was clean, then direct digestion of the mixture proved slightly more successful than analysis of gel-separated proteins. When an enriched Golgi complex from rat liver was separated by either 2-DE (135 spots) or 1-DE (77 bands) and in-gel digested followed by LC-MS/MS, significantly more proteins were identified from the 1-DE separation.

David Arnott (Genentech) described the proteomics component of Genentech's Secreted Protein Discovery Initiative, which also includes genomic, signal trap, expression, and functional analysis. Arnott evaluated three methods to identify proteins secreted from human umbilical microvascular endothelial cells (HUMECs) into 60 ml of serum-free media; 2-DE and 1-DE (with/without staining) followed by in-gel digestion, and direct digestion of the entire protein mixture. Digests were analyzed using the microcapillary system described above. Interestingly, direct digestion followed by data-dependent LC-MS/MS identified the most proteins, but all three methods were complementary in their hands (21 proteins identified by all three methods but no completely novel gene products).

Analysis of serum fractionated using the Cohn pH/ethanol precipitation protocol followed by digestion of the entire fraction prior to data-dependent LC-MS/MS was described by Karl Clauser (Millennium Pharmaceuticals) in the context of the studies of differences between wild type and ApoE −/− mice. Clauser also presented the bioinformatics flow for data handling, which utilizes a variant of the publicly available MS-Tag (**http://prospector.ucsf.edu/**) for protein identification and a de novo sequence interpretation program referred to as SHERENGA (2). Their stated aim is for searching to keep up with the LC-MS/MS. They have also been experimenting with the IEX ion-exchange protocol developed by Andy Link (7) as a means of decreasing the complexity of the sample and reducing the number of singly charged and highly charged ions as these are less likely to be identified. In one IEX

Physiological Genomics

fraction, 87 plasma proteins were identified in a single run compared with 66 from an unfractionated sample.

Scott Patterson (Amgen) described Amgen's proteomics efforts, now in the third year. They are employing data-dependent LC-MS/MS of complex protein mixture digests. The stated aim is to reduce the complexity such that in an ideal situation only one peptide for each protein in the mixture is fragmented during LC-MS/MS. To achieve this aim, various affinity methods can be employed, and the use of cysteinyl peptide capture using either thiopropyl Sepharose or a biotin alkylating reagent, *N*-[6-(biotinamido)hexyl]-3′-(2′-pyridyldithio)-propionamide (HPDP-biotin), was described (13). The former was used in a large-scale analysis of urinary proteins where digestions of the unfractionated starting material, albumin/IgG depleted, or cysteinyl peptide captured or noncaptured were analyzed. The samples were analyzed with replicate LC-MS/MS runs using narrow mass ranges for ion selection for each run, thereby increasing the number of unique spectra selected for fragmentation. This analysis resulted in the identification of over 200 proteins, including a number of uncharacterized nucleotide sequences (e.g., ESTs). Smaller scale analyses are described above, in one case [soluble intermembrane proteins (SIMPs)] utilizing cysteinyl peptide capture to identify more proteins than with no fractionation. Data handling for this high-throughput effort was also briefly described. A number of the fractions being analyzed have some of the same components; therefore, to enhance the identification process, spectral matching of the database (>5 million spectra) is performed. This links identical spectra and therefore reduces the redundancy associated with re-searching already identified spectra.

Quantitative analysis of two samples without electrophoresis. MALDI-MS, using the surface enhanced laser desorption ionization (SELDI) surface, to search for disease markers in biological fluids was presented by Scot Weinbeger (Ciphergen Biosystems). In this approach, defined chemical/biochemical surfaces are utilized to allow fractionation of proteins from biological fluids in a reproducible manner. This reproducibility allows comparisons between different samples to be made. Weinberger described the search for markers of benign prostatic hyperplasia that, like prostate cancer, displays elevated prostate specific antigen (PSA) levels. The fraction exhibiting a difference between these samples was able to be enzymatically digested, and a number of peptides were generated. These were able to be fragmented using the MALDI Qq-TOF of Standing, described above. It appears as though there is a difference in the relative level of a seminogelin fragment between these two diseases, providing a potential differential marker. The method is sensitive but apparently limited to analysis of proteins less than about 20 kDa (a range not well characterized by 2-DE).

A combination gel/MS approach referred to as a "virtual 2-D gel" was presented by Phil Andrews (Univ. of Michigan). In this approach, proteins are separated by charge using thin-layer isoelectric focusing (IEF), and this gel is then subjected to MALDI-MS. By rastering through the entire IEF gel, a composite display of all acquired MALDI-MS spectra can be generated (hence, the virtual 2-DE). Such analyses would provide very accurate mass measurements, greatly assisting in posttranslational modification analyses as well as potentially quantitation.

Karl Clauser (Millennium Pharmaceuticals) described their efforts at utilizing already existing LC-MS/MS data to attempt to gain some quantitative/qualitative information as to differences between samples. Differences in serum protein levels between wild-type and ApoE −/− mice have been examined using this approach, which compares the MS ion current from peptides identified between LC-MS/MS runs of each sample. Comparison between runs is a difficult task, but data suggested that there is sufficient confidence to state a significant difference if there is a difference of a factor of 3 between some components of the samples.

An LC-MS/MS-based system was described by Steve Gygi (Univ. of Washington) for quantitative analysis of complex mixtures. The technology is referred to as isotope-coded affinity tag (ICAT) (6). The ICAT reagent described here is composed of three units: an affinity reagent (biotin), a linker region (one of two forms), and a reactive group (a thiol-specific reagent, iodoacetic acid). The linker region encodes the mass difference, with the light version having 8 hydrogens and the heavy version having 8 deuteriums. Thus the mass difference is 8 mass units (doubly charged ions will have an m/z difference of 4). Following reduction and alkylation of each of the two protein samples with one of the two reagents, the two samples can then be mixed together. All subsequent manipulations are performed as a mixture, culminating in tryptic digestion of the complex sample and capture of the cysteinyl peptides on avidin. The bound peptides are released and analyzed by LC-MS/MS, revealing paired signals of peptides. Calculation of areas under the peak for each paired ion from the LC-MS data provides an accurate record of the relative quantities of the proteins from each starting sample. The MS/MS spectra allow identification of the peptides. The approach was elegantly demonstrated with yeast grown on either galactose-containing media or ethanol-containing media. Proteins expected to be differentially regulated were observed, and, highlighting the advantages of analysis at the protein level as opposed to the mRNA level (e.g., microarray), alcohol dehydrogenase-1 (ADH1) was found to be oppositely regulated (as expected) to ADH2, to which it is 95% homologous. This is a very promising approach for quantitative analysis of complex protein mixtures.

A number of interesting posters were also presented at the meeting, and some of the presenters were given the opportunity to "advertise" their posters. These dealt with the same range of subjects presented in the oral sessions.

Conclusion. The organizers Ruedi Aebersold and John Stults brought together an excellent program for this meeting, with essentially all major laboratories in

this field being represented. The field has grown enormously over the past few years, and advancements presented at this meeting indicate an optimistic view of the future for proteomics. This very successful meeting provided the 162 attendees with the state-of-the-art in mass spectrometry and proteomics.

Address for reprint requests and other correspondence: S. D. Patterson, Biochemistry, Amgen Inc., One Amgen Center Drive, MS 14-2-E, Thousand Oaks, CA 91320-1789 (E-mail: spatters@amgen.com).

REFERENCES

1. **Carr SA and Annan RS.** Overview of peptide and protein analysis by mass spectrometry. In: *Current Protocols in Molecular Biology*, edited by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K. New York: Wiley, 1997, p. 10.21.1–10.21.27.
2. **Dancik V, Addona TA, Clauser KR, Vath JE, and Pevzner PA.** De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol* 6: 327–342, 1999.
3. **Eng JK, McCormack AL, and Yates JR, III.** An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5: 976–989, 1994.
4. **Eriksson J, Chait BT, and Fenyo D.** A statistical basis for testing the significance of mass spectrometric protein identification results. *Anal Chem* In press.
5. **Gottschalk A, Neubauer G, Banroques J, Mann M, Luhrmann R, and Fabrizio P.** Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J* 18: 4535–4548, 1999.
6. **Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, and Aebersold R.** Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17: 994–999, 1999.
7. **Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, and Yates JR, III.** Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17: 676–682, 1999.
8. **Mintz PJ, Patterson SD, Neuwald AF, Spahr CS, and Spector DL.** Purification and biochemical characterization of interchromatin granule clusters. *EMBO J* 18: 4308–4320, 1999.
9. **Morris HR, Paxton T, Langhorne J, Berg M, Bordoli RS, Hoyes J, and Bateman RH.** High sensitivity collisionally-activated decomposition tandem mass flight mass spectrometer. *Rapid Commun Mass Spectrom* 10: 889–896, 1996.
10. **Morris HR, Paxton T, Panico M, McDowell R, and Dell A.** A novel geometry mass spectrometer, the Q-TOF, for low femtomole/attomole-range biopolymer sequencing. *J Protein Chem* 16: 469–479, 1997.
11. **Patterson SD.** Protein identification and characterization by mass spectrometry. In: *Current Protocols in Molecular Biology*, edited by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K. New York: Wiley, 1998, p. 10.22.1–10.22.24.
12. **Patterson SD and Aebersold R.** Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 16: 1791–1814, 1995.
13. **Patterson SD, Spahr CS, Daugas E, Susin SA, Irinopoulou T, Koehler C, and Kroemer G.** Mass spectrometric identification of proteins released from mitochondria undergoing permeability transition. *Cell Death Differ* In press.
14. **Qiu YC, Benet LZ, and Burlingame AL.** Identification of the hepatic protein targets of reactive metabolites of acetaminophen in vivo in mice using two-dimensional gel electrophoresis and mass spectrometry. *J Biol Chem* 273: 17940–17953, 1998.
15. **Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, and Seraphin B.** A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17: 1030–1032, 1999.
16. **Vogl T, Roth J, Sorg C, Hillenkamp F, and Strupat K.** Calcium-induced noncovalently linked tetramers of MRP8 and MRP14 detected by ultraviolet matrix-assisted laser desorption/ionization mass spectrometry. *J Am Soc Mass Spectrom* 10: 1124–1130, 1999.
17. **Wilkins MR, Sanchez J-C, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, and Williams KL.** Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13: 19–50, 1995.
18. **Wilm M, Neubauer G, and Bachi A.** Automatic de novo sequencing of proteins by differential scanning with a quadrupole time of flight instrument. *Proc 47th ASMS Conf Mass Spectrom Allied Topics Dallas June 13–17 1999*, p. 1541.
19. **Wilm MS and Mann M.** Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int J Mass Spectrom Ion Processes* 136: 167–180, 1994.

Mass spectrometry and proteomics

SCOTT D. PATTERSON

Physiol. Genomics 2:59-65, 2000.

Physiological Genomics publishes results of a wide variety of studies from human and from informative model systems with techniques linking genes and pathways to physiology, from prokaryotes to eukaryotes. It is published quarterly in January, April, July, and October by the American Physiological Society, 9650 Rockville Pike, Bethesda MD 20814-3991. ISSN: 1094-8341, ESSN: 1531-2267. Visit our website at http://www.the-aps.org/.

[CANCER RESEARCH (SUPPL.) 52, 2711s–2718s, May 1, 1992]

2132.096 Examiner copy

Considerations in Bringing a Cancer Biomarker to Clinical Application[1]

reference # 4

Melvyn S. Tockman, Prabodh K. Gupta, Norman J. Pressman, and James L. Mulshine

Department of Environmental Health Sciences, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205 [M. S. T.]; Department of Pathology and Laboratory Medicine, The University of Pennsylvania Medical Center, Philadelphia, Pennsylvania 19104 [P. K. G.]; Cell Systems International, Inc., Rockland, Delaware 19732 [N. J. P.]; and Biomarkers and Prevention Research Branch, National Cancer Institute, Bethesda, Maryland 20814 [J. L. M.]

Abstract

Specific challenges face our application of emerging biomarkers to early lung cancer detection. These challenges might be considered frontiers to be bridged between established biomedical disciplines, requiring expertise often beyond the range of individual investigators. Cross-disciplinary research already has led to new appreciation of the mechanisms which underlie the phenotypic expression of the transformed cell and places within our grasp the tools which might lead to successful early lung cancer detection. Prior to the successful application of newly described markers, further cross-disciplinary research must (*a*) refine the selection of biologically appropriate markers, (*b*) validate such markers against acknowledged disease end points, (*c*) establish quantitative criteria for marker presence/absence, and (*d*) confirm marker predictive value in prospective population trials.

During the 1960s and 1970s, the only clinical marker available to detect pulmonary neoplastic changes was the recognition of morphological atypia in exfoliated epithelial cells by light microscopy (1). We now know that cytomorphological criteria alone are not sufficiently sensitive for lung cancer screening. The three National Cancer Institute-sponsored clinical trials (at Johns Hopkins University, Memorial Sloan-Kettering Hospital, and the Mayo Clinic), have demonstrated, among 30,000 high-risk participants, that chest radiography and sputum cytology can detect presymptomatic, earlier-stage carcinoma, particularly carcinoma of the squamous cell type (2–5). Higher resectability and survival rates among the study groups than in the controls did not translate, unfortunately, into lowered (overall) lung cancer mortality. Less than 10% of lung cancers in the early lung cancer detection trial were detectable only by routine sputum cytology. Length-biased sampling, lead-time bias, and misclassification, in addition to failures of detection and of intervention, contributed to the lack of improvement in mortality rates (6–8).

In the intervening years, with the explosive interest in tumor biology, new tools have emerged with a greater potential to identify markers of neoplasia in the sputum. As summarized in Table 1 (adapted from Ref. 8), a variety of biological tools could be used for evaluation of carcinogenesis in shed bronchial cells. Which of these targets ultimately might function most effectively as a screening tool is still a matter for speculation. Monoclonal antibody recognition of tumor-associated antigens has progressed furthest toward application as a lung cancer biomarker and will illustrate this discussion. By closely following the biology of lung carcinogenesis, other rationally developed diagnostic tools can potentially detect the process of carcinogenesis before the clinical onset of cancer. A wealth of published information, including our own experience (9), is relevant in attempting to understand and organize the complex issues involved in bring a biomarker into applications for preventive approaches to lung cancer. This report attempts to survey and evaluate salient points in this process.

[1] Presented at the NCI Workshop "Investigational Strategies for Detection and Intervention in Early Lung Cancer," April 21–24, 1991, Annapolis, MD. This research was supported in part by a Collaborative Research and Development Agreement between the National Cancer Institute, the Johns Hopkins Medical Institutions, the University of Pennsylvania, and Abbott Laboratories.

Biomarker Identification and Selection

The process of bringing these (or other) biomarkers from the laboratory to application in human populations requires new insight at four frontiers, *i.e.*, research domains which overlap between traditional disciplines. The domain of biomarker identification falls primarily to the tumor biologist who, in seeking to apply a marker for early detection, finds himself confronted with three fundamental issues: (*a*) to provide a clear definition of the end point for which the putative index is a marker; (*b*) to identify the type of clinical specimen from which the marker can be measured; and (*c*) to establish an expected (*i.e.*, normal or background) range of marker variability.

End Point Definition. The new paradigm of carcinogenesis presents a multistage model which has potential genetic or epigenetic markers at each stage. Tumor initiation marks the genetic change from a normal to an initiated cell. Oncogene activation and/or inactivation of tumor suppressor genes may emerge as markers of this earliest stage of carcinogenesis (although these also may occur later at the stage of tumor progression) (10). Activation of six families of oncogenes, *ras*, *raf*, *jun*, *fur*, *neu*, and *myc*, have been associated with human lung cancer (10), albeit the activation sequence defining which oncogene may be involved with the earliest stage of carcinogenesis has not been established. Ki-*ras* expression has been frequentl detected in bronchial adenocarcinoma tumor tissue (11, 12) associated with shortened survival for both early stage (13) an advanced disease (14). Elevation of c-*myc* expression has bee associated with growth deregulation and loss of terminal differ entiation in squamous cell and in small cell tumors (15). Het erogeneity for c-*myc* amplification and rearrangement of *my* family oncogenes suggest that activation of these oncogen may occur later in carcinogenesis, during tumor progressio (16).

Loss of transcription factors for suppressor genes may b come markers of human lung cancer. Using probes to dete allelic deletion of specific chromosomal regions by restricti fragment length polymorphisms, Yokota *et al.* (17) has fou frequent loss of heterozygosity (expression of only a sin allele) in human small cell lung tumors on chromosomes (100%), 13q (91%), and 17p (100%). Loss of heterozygosity chromosome 3p was also detected in the tumors of 83% adenocarcinoma patients. Miura *et al.* (18) found the karyoty in non-small cell lung cancer to be very complex, but recurr loss of 17p, 3p, and 11p (in 67, 57, and 48% of cases, resp tively) suggests "hot spots" for genetic alteration in lung can Other candidate regions with breakpoints indicating poten recessive oncogenes include 1q, 3q, 5p, 7p, 16q24, and (19).

Oncogene activation may alter the metabolic balance betw cell growth and differentiation (10). By encoding "transc tion" proteins, oncogenes activate other key genes to code growth deregulation. The shift in the balance from cell (te nal) differentiation to growth marks the selective clonal ex sion characteristic of tumor promotion. Two critical " response" transcription factors, *fos* and *jun*, seem to be activ

Table 1 *Potential targets in bronchial fluids and/or sputum for early lung cancer detection*[a]

Differentiation markers (*e.g.*, glycolipid expression)
Specific tumor products (*e.g.*, mucins, matrix proteins, surfactant)
DNA ploidy
Polyamines
Nucleosides
Growth factors
Oncogenes or oncogene products
Cytogenetic changes
Specific chromosomal deletions or rearrangements
DNA repair enzymes
DNA adducts

[a] Adapted from Ref. 8.

whenever mammalian cells respond to peptide growth factors (20, 21). Bombesin, for example, a peptide growth factor released by pulmonary neuroendocrine cells, has been shown to induce growth and maturation of human fetal lung in organ reculture (22). A functional membrane-associated bombesin receptor recently has been isolated from human small cell lung carcinoma (NCI-H345) cells (23), and bombesin-like peptides have been found in the bronchial lavage fluid of asymptomatic cigarette smokers (24). Thus markers of growth factor expression, insofar as they reflect oncogene activation, may also hold promise for the detection of early (preneoplastic) lung cancer.

The biology of gene transcription and signal transduction leads us to suspect that cytoplasmic and cell surface products of activated oncogenes will be far easier to detect (*i.e.*, more frequent binding sites) than would the detection of specific allelic polymorphism. Expression of tumor-associated antigens may be the only detectable signal if cancer is to be recognized before tumor tissue can be clinically detected and biopsied. Immunization of mice with lung cancer-associated antigens led to the development of antibody-producing hybridomas from which monoclonals were selected by their preferential reactivity with tumor cells over normal cells (25, 26). The two murine antibodies selected for our analysis exhibited the best reactivity against SCC[2] and NSCC cell lines and clinical specimens (27). Hakomori has noted that there is no unique, "tumor-specific" chemical structure responsible for the specificity of tumor-associated antigens (28), although Hakomori (29) and others (30) have shown that many of these differentiation markers are defined by carbohydrate antigens. An example Hakomori (28) cites is SSEA-I, a tightly regulated developmental marker with limited expression in mature tissues. While highly expressed at the tumor cell surface, these tumor-associated antigens are absent in progenitor cells and show limited expression on other normal tissues. Monoclonal antibodies may recognize either this high antigen density or a specific conformation (epitope) induced by the high density. It is believed that this antibody recognition of a density-specific conformation defines tumor specificity (28). Expression of such tumor-associated antigens could follow neoplastic transformation of the pulmonary epithelial cells. Activated oncogenes coding for transcription factors may permit the reexpression of fetal differentiation markers such as the carbohydrate structure SSEA-I. Alternatively, SSEA-I epitope expression might be the result of unspecified posttranslational enzyme modification (29). Posttranslational phosphorylation, for example, has been suggested as a reversible mechanism by which to modify the activity of cellular proteins (31).

[2] The abbreviations used are: SCC, small cell carcinoma; NSCC, non-small cell carcinoma; SSEA-I, stage-specific embryonal antigen I; JHLP, Johns Hopkins Lung Project; DAB, diaminobenzidine.

In our recent report of successful immunocytochemical staining of sputum samples for early lung cancer detection (27) we extend the directions suggested by Saccomanno *et al.* (1) and Hakomori (28, 29). Tockman *et al.* (27) used two lung cancer-associated monoclonal antibodies to stain sputum specimens acquired in the course of the JHLP component of the National Cancer Institute-sponsored early lung cancer detection trial. Approximately one-half (5,226) of the 10,386 community-dwelling, high-risk individuals (males, age >45 years, and currently smokers of ≥1 pack/day) had been randomly allocated to receive cytological plus radiographic (dual) screening. Moderate atypia was found on one or more specimens of 626 (12%) of these individuals. The first atypical and all subsequent specimens of these individuals were preserved in Saccomanno's preservative (50% alcohol and 2% Carbowax 1540) at room temperature. The first morphological atypical specimens of these 626 JHLP participants were divided into four groups (Table 2). Two of the groups consisted of specimens that demonstrated only moderate atypia; most (n = 537, 86%) of these participants never went on to lung cancer (group I). However, it is important to observe that from the 40 (6.4%) individuals who did progress to lung cancer (group II), all four major lung cancer cell types eventually arose. Groups III and IV consisted of specimens with marked atypia on at least two occasions. Three individuals (group III) never developed lung cancer. The majority of those with marked atypia (group IV) progressed to cancer; all were NSCC, and the majority were of the epidermoid cell type. These observations suggest that cells exfoliated at the stage of moderate atypia may be a morphological correlate to a neoplastic stem cell capable of differentiation into all four major cell types of lung cancer.

Murine monoclonal antibodies 624H12 and 703D4, which bind to a glycolipid antigen of small cell (32) and protein antigen of non-small cell lung cancer (33), respectively, were applied to the alcohol-fixed (preserved) sputum specimens collected from JHLP participants with at least moderate atypia who later developed lung cancer and from controls who did not. Immunostaining was applied to the earliest preserved specimens, using a double-bridge immunoperoxidase technique with a biotinylated DAB chromogen (34). Specimens from individuals who ultimately developed lung cancer stained wit a sensitivity of 91%, 2 years (on average) before the clinica appearance of neoplasia. Specificity was 88% among specimen

Table 2 *Allocation of JHLP participants with stored sputum by atypia grade an cell type of eventual lung cancer*[a]

Atypia severity/cancer development	*N*	%
Moderate atypia	626	100
Group I atypia < marked (×2)	537	86
No lung cancer		
Group II atypia < marked (×2)	40	6.4
Squamous	12	
Small cell	9	
Adeno-	7	
Large cell	8	
Other, mixed	4	
Group III atypia > marked (×2)	3	0.5
No lung cancer		
Group IV atypia > marked (×2)	46	7.4
Squamous	41	
Adeno	3	
Large cell	2	

[a] Adapted from Ref. 27.

from individuals who remained free of lung cancer (Table 3) (27).

Other investigators (25, 35) have shown that oncofetal antigen expression is not limited to lung cancer. The epitope of SSEA-I, a fucosylated ceramide pentasaccharide (lacto-*N*-fucopentaose III), was found to be immunodominant for mouse-derived monoclonal antibodies for a variety of epithelial malignancies. The contrasting expression of lacto type II antigens in normal and malignant epithelial tissues recently has been mapped (29, 36). Complexity (numbers of fucosyl repeats) of glycosphingolipid antigen expression is correlated with nuclear differentiation and with survival (30). If cancer is considered a retroversion of ontogeny, then the study of such differentiation markers may provide an extraordinarily useful map of the early steps of carcinogenesis. Differentiation markers could also provide useful prognostic clues of tumor growth and response to therapy.

The emphasis in this section upon markers of the early stages of carcinogenesis was intentional. Evidence of a transformed genome, by expression of tumor-associated antigens, oncofetal growth factors, or specific chromosomal deletions, has clear biological plausibility as a marker of preclinical lung cancer. In contrast, nonspecific markers of gene damage, *e.g.*, DNA adducts (37), micronuclei (38, 39), and sister chromatid exchanges (40), in the absence of evidence of gene transformation, might be considered markers of exposure or susceptibility. Susceptibility markers act as effect modifiers, whereby subgroups of otherwise similar individuals show an enhanced disease response (probability of disease) (41). Marker selection depends upon the investigator's hypothesis, of course, and both markers of individual early disease and markers of enhanced group risk have a role. If markers of early disease can be validated, then markers of susceptibility can be useful for population screening. This dual-phase strategy whereby a simple, highly sensitive, but not necessarily specific initial screen might be followed by a more specific, confirmatory test has been suggested recently as most suitable for the vast populations at risk (42).

Type of Tissue Specimen. Appreciation that early lung cancer invades the pulmonary parenchyma from the bronchial epithelium led Saccomanno *et al.* (1) to suggest that exfoliated epithelial cells recovered in the sputum may provide intermediate end points of developing lung cancer. Although cytomorphology-based lung cancer screening failed to reduce overall lung cancer mortality in the National Cancer Institute-sponsored Early Lung Cancer detection trial (7), the original concept of early lung cancer recognition through detection of exfoliated cell markers remains intact. Enthusiasm for detection of markers on exfoliated airway epithelial cells stems in part from the belief that alternative (blood-borne) markers become detectable only after bronchogenic cancer crosses the basement membrane and invades the pulmonary parenchyma and vascular system, *i.e.*, becomes a systemic (not localized) disease. On theoretical grounds, one would not expect blood-borne markers to permit identification of localized lung cancer. On empirical grounds, a variety of tumor markers in serum have been evaluated for use to detect early cancer, including carcinoembryonic antigen, CA-125, and many others. None of these has been accepted for clinical application in this setting (43). Cellular elements of peripheral blood also have been examined for lung cancer markers. Peripheral blood lymphocyte DNA has been extracted and digested with restriction endonuclease (*Msp*I) to evaluate the association of restriction fragment length polymorphisms of the P450IA1 gene with lung cancer (44). Although the frequency of a homozygous rare allele of P450IA1 gene was 3-fold higher among lung cancer patients than among healthy controls, no information is available regarding this marker with the early (preclinical) stages of carcinogenesis. Studies of genetic changes in peripheral blood lymphocytes may eventually indicate host exposure/susceptibility but are not expected to detect the early stages of lung carcinogenesis.

Table 3 *Result of double-bridge immunoperoxidase staining of monoclonal antibody surface markers applied to first atypical sputum specimen*[a]

	Lung cancer	No lung cancer	Total
Satisfactory			
Stain +	20	5	25
Stain −	2	35	37
Subtotal	22	40	62
Unsatisfactory	4	3	7
Total	26	43	69

[a] Adapted from Ref. 27. Note: Sensitivity = 91%; specificity = 88%; odds ratio = 70; 95% confidence interval = 10.46–297.8; $P < 1 \times 10^{-4}$. Or atypical specimen approximately 2 years in advance of clinical cancer.

While cytologists have long been aware that multiple sputum specimens may adequately sample the airway epithelium at risk of cancer, particularly for centrally located lesions (45), it is not yet known conclusively whether a focus of preneoplasia may be adequately represented in the sputum. Nevertheless, there is strong suggestive evidence that preneoplastic cell markers are detectable in the sputum. Frost *et al.* (6) have shown that the greater the degree of atypical (preneoplastic) sputum cytomorphology the more likely is the subsequent diagnosis of lung cancer. The 12% of the dual-screened participants in the Johns Hopkins early detection trial who produced moderate (or more severe) atypia accounted for more than one-third (86 of 233; 37%) of the lung cancers which developed over the subsequent 5 to 8 years (27). Further evidence for the extent of the affected epithelium comes from studies of the entire bronchial tree (from the surgical margin of resection to the ends of the subsegmental bronchi) from patients with early (*in situ* and microinvasive) lung cancer which have shown large areas of neoplastic epithelium surrounding the neoplastic focus (46, 47). Lippman *et al.* (48) has adopted the concept of "field cancerization" of Slaughter *et al.* (49) to describe the widespread, multifocal morphological change observed in the airway epithelium assaulted by inhaled carcinogens (*i.e.*, tobacco smoke). The ease of obtaining a sufficient number of cells which express tumor-associated antigen molecules for immunoprobe detection in sputum makes this marker/medium combination attractive. With future refinements, molecular evidence of specific genome transformation may come from biological material provided from sputum or endoscopic biopsy.

Difference from Background. Every potential marker has its inherent unreliability, its inevitable lack of constancy, when the measure is applied repeatedly to the same individual (50). If a biological change is to be an indicator of disease, it must produce a recognizable departure from "normal"; the "change" produces a difference from the mean or usual value by an amount greater than is likely due to random or expected variability. Having established that a biological index may be useful for recognizing the early stages of carcinogenesis, the interaction of the biologist with a statistician/epidemiologist is recommended.

Peto *et al.* (51) have responded to a report that the inheritance of rare hypervariable alleles at the Ha-*ras*-1 locus is associated with a predisposition to human cancer, with five general guidelines. In brief, these investigators recommend that: (*a*) the initial

analysis of association between marker frequency and cancer should be made upon the data of the entire study population, not by subgroup analysis; (*b*) consider bias as a possible explanation for a positive association; (*c*) if bias can be excluded, then inspect for a data-derived hypothesis to account for the variation; (*d*) test the data-derived hypothesis on a fresh cohort to support the validity of the marker hypothesis before its publication; and (*e*) assure the biological plausibility of the data-derived hypothesis.

Biomarker Validation against Acknowledged Disease End Points. Following selection of a biomarker, the sensitivity and specificity of label-epitope binding in premalignant specimens must be validated to a known (histology/cytology-confirmed) cancer outcome. This domain, therefore, is the province of the pathologist, with some communication with the statistician/epidemiologist.

Prior to our testing of monoclonal antibody in patient sputa, monoclonal antibodies were selected based upon binding in tissue culture and histological sections of tumor and normal lung (27). The precision of epitope localization which results from optimal monoclonal antibody selection was further enhanced by modifying the avidin-biotin-peroxidase complex immunostain by the method of Gupta *et al.* (34). This modification entails the addition of a second biotinylated antibody-avidin-biotin-peroxidase complex reagent layer. The species-specific secondary immunoglobulin binds to the layers previously applied and increases the size of the lattice-like bridge between the antigen and the enzyme molecules that catalyze the staining reaction. This method has been shown to enhance immunostaining in both Saccomanno-fixed cytological and paraffin-embedded histological tissue.

The terms sensitivity and specificity indicate how well a particular biological change indicates a disease (*e.g.*, cancer) outcome. These indices are usually determined by applying the marker to specimens from one group of persons who have (or will develop) the disease and to specimens from another group who do not and then comparing the results. For our discussion, sensitivity refers to the proportion who stain positive among all those with/who develop cancer, while the specificity is the proportion who stain negative among all without/don't develop cancer (52). The more sensitive an indicator is for a disease and the more specific it is for that disease only, the better it functions as a test. In contrast to "predictive value" or "false positive rate," the sensitivity and specificity are constant for a given test even when different groups or populations are tested (53).

The essential element of the validation of an early detection marker is the ability to test the marker on clinical material obtained from subjects monitored in advance of clinical cancer and link those marker results with subsequent histological confirmation of disease. The experience gained through the National Cancer Institute early detection trial at Johns Hopkins again is instructive. That study provided an archived bank of sputum specimens with a record of the clinical course and long-term follow-up for the patients from whom the specimens were obtained (3–5). Clinical follow-up for an average of 8 years from specimen collection was available for all of the 626 individuals who showed moderate (or greater) atypia (27). Histological slides were obtained for almost every case from biopsy or autopsy to confirm the link between the intermediate end point and a standard, pathologically confirmed definition of a case. A reasonable (5- to 8-year) cancer-free follow-up period was required for each control. Other investigators at M. D. Anderson are presently preserving a bank of endoscopic samples from individuals with high risk of lung cancer for histological follow-up of subsequent cases and controls for validation of their marker studies (38). This irrefutable link between antecedent marker and subsequent acknowledged disease is the essence of a valid intermediate end point.

Quantitative Criteria for Marker Presence/Absence

The lack of a unique chemical structure for tumor-associated antigens signifies that a qualitative presence/absence criterion of marker binding is insufficiently specific. These circumstances require the development of rigorous quantitative criteria for positive/negative marker binding based upon the number of probe adherence sites per cell and the frequency of labeled cells per specimen. We have been engaged in studies to quantify immuno-labeled cell detection by characterizing the source and magnitude of the optical/electronic probe signal compared to all other sources of variation (the noise) (54). Noise may arise from technical variation in the specimen collection/preparation, from variation in the assay, from biological variation in the host (*e.g.*, in the degree of cytological atypia), and in the quantitation of marker uptake.

Automated cytology systems are able to augment human ability to detect and interpret biologically significant cellular and tissue changes (6, 55, 56). Automated instruments are capable of determining spectral characteristics of stained cellular proteins, DNA content, and ploidy (57–61). Commercially available integrated, optical microscope, and computer systems are available to enable the pathologist to recognize morphological and cytochemical markers for solid tumors of the bladder, breast, colon, lung, pancreas, prostate, and thyroid and for non-Hodgkins lymphoma (62–72). However, recognition of biomarkers for preneoplastic lesions represents a new departure for this technology. Greenberg *et al.* (73) have focused their initial studies upon automation of traditional morphometry of atypical cells in sputum specimens, leading these investigators to the development of a "cell atypia profile" which may prove useful as a marker of carcinogenesis if validated in clinical specimens such as those available for the present study.

Our preliminary analyses have focused upon quantitation of three image-derived properties (spectral signature, optical texture, and morphology) of both labeled and unlabeled malignant cells. Probe characterization (*i.e.*, marker recognition studies) has been accomplished initially by performing spectral analyses of neoplastic (SCC and NSCC) cell lines, prepared using the standard Saccomanno technique for sputum cytology, followed by DAB immunostaining and methylene blue counterstaining of specimens which have been incubated with and without the primary antibody (for positive and negative controls, respectively). The morphometric (*e.g.*, size and shape parameters) and photometric (*e.g.*, texture and DNA content/distribution) features were analyzed with conventional univariate statistical techniques.

Spectral Signature. Transmission spectra were obtained over the visible spectrum (*i.e.*, 400 to 700 nm) for multiple cells/slide from each population. Log-absorbance (*i.e.*, log optical density) spectra were computed and used to estimate the variability of the spectral characteristics of the probe within slides (*i.e.*, intraslide variability) and between slides (*i.e.*, interslide variability) (Fig. 1). These spectral studies were performed to determine the optimal wavelength(s) to collect morphometric and densitometric data. The spectra of probe-positive and probe-negative specimens were measured on the Zeiss Axiomat

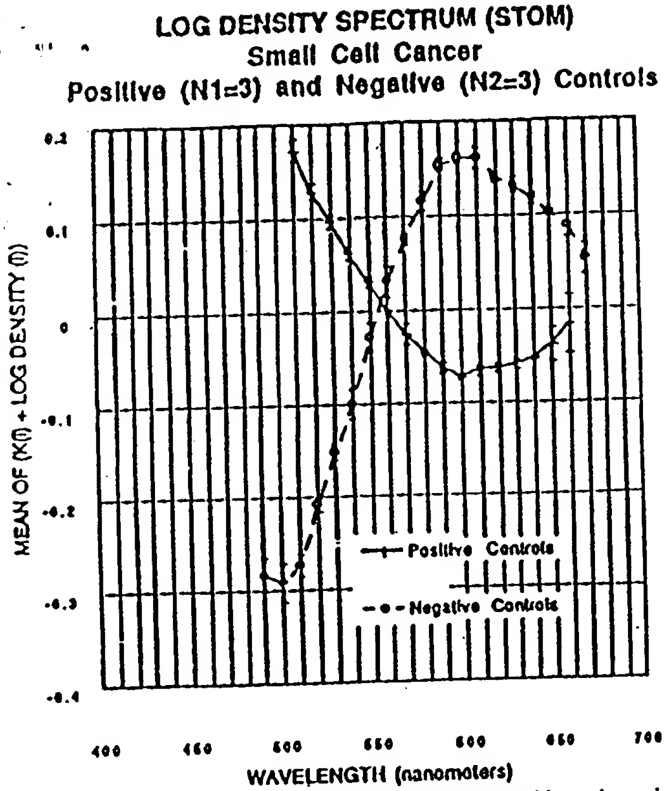


Fig. 1. Comparison of log optical density spectrum of positive and negative controls for SCC samples. These data show that the average log optical densities of the positive controls ($N = 3$) and the average of the negative controls ($N = 3$) are maximally different at frequencies of 510 and 600 nm. The presence of a stained DAB label on the positive controls renders them maximally optically dense at the shortest measured wavelengths. The methylene blue counterstain renders the negative controls maximally optically dense at approximately 600 nm. These plotted data are from measured spectra that have been normalized by additive constants, so that one cannot compare differences in absolute optical densities from these data.

microscope in the Frost Center Laboratory at the Johns Hopkins School of Hygiene. The log optical density spectra of probe-negative and probe-positive cells were compared. Results show that positive and negative cells are maximally different at approximately 510 nm and at 600 nm. These wavelengths maximally discriminate DAB-labeled cells from unlabeled (but counterstained) cells given the estimated variability associated with both the control specimens and their preparation methods. A spectral ratio parameter was tested successfully with respect to its discriminatory potential to separate control-negative cells from control-positive and sputum-derived cells. Narrow-band dual-wavelength optical scanning appears to be a powerful approach to discriminate probe from counter stain, according to this study.

Optical Texture. Optical texture was determined quantitatively by measuring statistical moments of the frequency distribution of optical densities within the cytoplasm of individual cells. The frequency distributions of cytoplasmic optical densities may be useful for cell class discrimination when the histograms are normalized for cytoplasmic area. Measurements of the variation of absorbance measurements within cells and cytoplasm show that the cells under investigation have variable optical densities (*i.e.*, are textured). A texture parameter (*i.e.*, short-run-length-emphasis) was measured for cells from each of the six classes under investigation. This single feature, when applied to the measurement of cytoplasmic texture, accurately discriminated each of the 6 control-negative cells from the total sample of 18 measured cells including an additional 6 control-positive cells and an additional 6 cells from the sputum of patients (Fig. 2).

Morphology. Morphometric studies of cell areas and shapes were conducted by measurement of nuclear area and total cell area and by analysis of Fourier coefficients of closed linear contours. Area measurements were evaluated after normalization to μm^2 units. Shape was quantified using simple parameters such as P**$2/A$ and by the Fourier coefficients of the closed linear contours that define the cell boundaries. Results showed that manually traced cell boundaries were reproducible compared to intercellular or internuclear area measurements, although the shape of manually traced nuclei varied more significantly from tracing to replicate tracing. Results also show that the shapes of the cells from the sputum cases sampled were more irregular than those of either positive or negative control cells. Fourier analysis was tested on these data and demonstrated a potential for discriminating irregularly shaped nuclei from those with more nearly round shapes.

Features with demonstrated discriminatory potential will be compared by correlation matrix analyses. Multivariate (*e.g.*, stepwise linear discrimination analysis) statistical techniques will be used to produce discriminant functions that combine relatively independent (*i.e.*, orthogonal) parameters. Training sets will be used to generate these discriminant functions and the functions will be tested on test data sets to determine their prognostic value in differentiating atypical cells from patients who subsequently developed lung cancer from those atypical cells from patients who did not.

Biomarker Confirmation in Population Specimens

While the overall sensitivity and specificity of sequential staining of replicate specimen slides with either of the two monoclonal antibodies are quite high (Table 3), it is apparent (Table 4) that the existing estimates of antibody sensitivity for

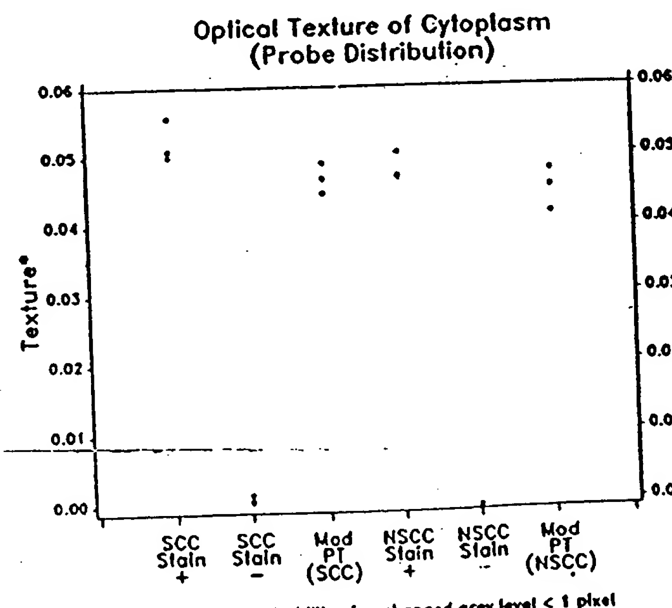


Fig. 2. Cytoplasmic texture as a function of cell type/stain group. Texture is more markedly reduced among negative controls than in either positive controls or patient cells which stained positively. This suggests that the texture of cytoplasm of the negative control cells is more finely varying than the texture of the cytoplasm of the cells which take up the DAB probe.

Table 4 *Monoclonal antibody sensitivity by lung cancer cell type*[a]

	Non-small cell antibody (703D4)			Small cell antibody (624H12)		
Cell type	N	Sensitivity (%)	95% CI[b]	N	Sensitivity (%)	95% CI
Adeno	5	60	14.7–94.7	5	0	0.00–0.05
Squamous	12	83	51.7–97.9	12	25	5.50–57.0
Large cell	4	50	6.80–93.2	4	0	0.00–0.04
Small cell	5	100	99.5–100	5	100	99.5–100

[a] Adapted from Ref. 27.
[b] CI, confidence interval.

Table 5 *Sample size required for selected values of sensitivity (with 95% confidence limits fixed at* $\pm\delta$*)*[a]

Sensitivity	$\delta = 4\%$	Sample size	$\delta = 6\%$	Sample size
99	95–100	24	93–100	11
95	91–99	114	89–100	50
91	87–94	197	85–97	87
80	76–84	384	74–86	171
70	66–74	504	64–76	224
60	56–64	576	54–66	256
50	46–54	600	44–56	266

[a] $\delta = Z^{*}SE_{Sensitivity}$ (one-half the range of confidence interval at each sensitivity).

each lung cancer cell type suffer from small numbers of observations. This is illustrated in Table 4 by the wide confidence interval around the sensitivity estimates.

Using the binomial distribution it is possible to calculate the sample size required to estimate the cell type subgroup sensitivity more accurately.

$$\text{Sample size} = \frac{(Z * \alpha/2)^2 * (p * q)}{\delta^2}$$

where $Z * \alpha/2$ is 1.96 upper percentage points of the standard normal distribution corresponding to the significance level $\alpha/2$ = 0.25, p is the proportion of true positive (in the calculation of sensitivity) or the proportion of true negative (in the calculation of specificity). q is $1 - p$, and δ is the specific difference between estimate and the "true value" (0.04 or 0.06 as in Table 5). It may be seen that the sample size required increases as the precision of the estimate increases [or the width of the estimated confidence interval decreases from $\delta = 6\%$ to $\delta = 4\%$; *e.g.*, for a sensitivity of 95%, the sample size must increase from 50 to 114; for a sensitivity of 99%, the sample size rises from 11 to 24 (Table 5)]. Thus, it is clear that the existing data are insufficient to accurately estimate cell type (subgroup)-specific sensitivity/specificity. Finally, as the risk of disease in a population falls, the positive predictive value of a test declines. Therefore, population application of even a valid test will be justified only after the predictive value of an early detection marker has been balanced against the population incidence of lung cancer. The development of such a strategy is the domain of the epidemiologist, who, through an ongoing dialogue with the biologist, may sequentially combine validated early detection markers to greatly enhance the accuracy of marker-based population screening.

We are currently engaged in a study to determine the validity (sensitivity, specificity, predictive value) of these monoclonal antibodies as markers of a new/continuing process of lung carcinogenesis in a population sufficiently large to provide for cell type (subgroup) analyses. Patients who have undergone successful resection for postsurgically staged (Stage I) lung cancer have a 5% annual incidence of developing a second primary lung cancer (74). Over a 3-year period, approximately 900 of these patients will be recruited to provide informed consent and undergo questionnaire interview, forced expiration, and sputum induction. One year of observation after the 3-year recruitment period will provide individual patient follow-up periods of 1–3 years. Diagnosis and treatment of second primary lung cancer will follow standard clinical practice. All lung cancer diagnoses and all causes of death will be clinically confirmed with pathology review. The sputum specimens will be stained by routine (Papanicolaou) and immunological methods to determine the validity of morphology and tumor-associated antigen detection independently and together for recognition of second primary lung cancer. Specimens will be preserved for the validation of new and refined markers of the earliest stages of carcinogenesis.

Summary

Developments in tumor biology in general and monoclonal antibody recognition of tumor associated antigens in particular hold great promise for detection of the stages of carcinogenesis well in advance of clinical cancer. However, prior to our application of emerging biomarkers to population-based early lung cancer detection, we face a series of collaborative research challenges. These challenges might be considered frontiers to be bridged between established biomedical disciplines, requiring expertise often beyond the range of individual investigators. Prior to the successful application of newly described markers, further cross-disciplinary research must (*a*) refine the selection of biologically appropriate markers: selection of valid markers of the initiated cell which appear in clinically accessible material is fundamental; (*b*) validate such markers against acknowledged disease end points: the use of specimen banks to validate existing and anticipated markers will be essential to rapid progress in the development of intermediate end points; (*c*) establish quantitative criteria for marker presence/absence: the absence of tumor-specific end points for tumor-associated antigens requires, at least for this class of markers, that quantitative criteria be established; and (*d*) confirm marker predictive value in prospective population trials: marker predictive value must be demonstrated in populations trials allowing for the anticipated effect of population disease risk upon test validation.

As progress in elucidation of the biology of early carcinogenesis is integrated with biomarker validation, new clinical applications will require an ongoing communication across traditional biomedical disciplines. Expansion of collaborative research leading to rational, practical application of validated cancer biomarkers may evolve, in turn, into a clinical field of early detection and prevention.

References

1. Saccomanno, G., Archer, V. E., Auerbach, O., Saunders, R. P., and Brennan, L. M. Development of carcinoma of the lung as reflected in exfoliated cells. Cancer (Phila.), *33:* 256–270, 1974.
2. Frost, J. K., Ball, W. C., Jr., Levin, M. L., Tockman, M. S., Baker, R. R., Carter, D., Eggleston, J. C., Erozan, Y. S., Gupta, P. K., Khouri, N. F., *et al.* Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study. Am. Rev. Respir. Dis., *130:* 549–554, 1984.
3. Frost, J. K., Ball, W. C., Jr., Levin, M. L., and Tockman, M. S. Final report: lung cancer control, detection and therapy, phase II. NCI Publication No. (PHS) N01-CN-45037. Washington, DC: National Cancer Institute, 1984.
4. Tockman, M. S., Levin, M. L., Frost, J. K., Ball, W. C., Jr., Stitik, F. P., and Marsh, B. R. Screening and detection of lung cancer. *In:* J. Aisner (ed.), Lung Cancer, pp. 25–40. New York: Churchill Livingstone, 1985.
5. Stitik, F. P., Tockman, M. S., and Khoury, N. F. Chest radiology. *In:* A. B.

Miller (ed.), Screening for Cancer, pp. 165–199. San Diego, CA: Academic Press, 1985.

6. Frost, J. K., Ball, W. C., Jr., Levin, M. L., Tockman, M. S., Erozan, Y. S., Gupta, P. K., Eggleston, J. C., Pressman, N. J., Donithan, M. P., and Kimball, A. W. Sputum cytopathology: use and potential in monitoring the workplace environment by screening for biological effects of exposure. J. Occup. Med., *28:* 692–703, 1986.
7. Tockman, M. S. Survival and mortality from lung cancer in a screened population — the Johns Hopkins Study. Chest, *89:* 324S–325S, 1986.
8. Mulshine, J. L., Tockman, M. S., and Smart, C. R. Considerations in the development of lung cancer screening tools. J. Natl. Cancer Inst., *81:* 900–906, 1989.
9. Mulshine, J. L., Linnoila, R. I., Jensen, S. M., Magnani, J. L., Tockman, M. S., Gupta, P. K., Scott, F. S., Avis, I., Quinn, K., Birrer, M. J., Treston, A. M., and Cuttitta, F. Rational targets for the early detection of lung cancer. J. Natl. Cancer Inst., in press, 1992.
10. Harris, O. C., Reddel, R., Modali, R., Lehman, T. A., Iman, D., McMenamin, M., Sugimura, H., Weston, A., and Pfeifer, A. Oncogenes and tumor suppressor genes involved in human lung carcinogenesis. Basic Life Sci., *53:* 363–379, 1990.
11. Pulciani, S., Santos, E., Lauver, A. V., Long, L. K., Aaronson, S. A., and Barbacid, M. Oncogenes in solid human tumours. Nature (Lond.), *300:* 539–542, 1982.
12. Rodenhuis, S., van de Wetering, M. L., Mooi, W. J., Evers, S. G., van Zandwijk, N., and Bos, J. L. Mutational activation of the K-*ras* oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. N. Engl. J. Med., *317:* 929–935, 1987.
13. Slebos, R. J., Kibbelaar, R. E., Dalesio, O., Kooistra, A., Stam, J., Meijer, C. J., Wagenaar, S. S., Vanderschueren, R. G., van Zandwijk, N., Mooi, W. J., *et al.* K-*ras* oncogene activation as a prognostic marker in adenocarcinoma of the lung. N. Engl. J. Med., *323:* 561–565, 1990.
14. Mitsudomi, T., Steinberg, S. M., Oie, H. K., Mulshine, J. L., Phelps, R., Viallet, J., Pass, H., Minna, J. D., and Gazdar, A. F. *ras* gene mutations in non-small cell lung cancers are associated with shortened survival irrespective of treatment intent. Cancer Res., *51:* 4999–5002, 1991.
15. Birrer, M. J., Raveh, L., Dosaka, H., and Segal, S. A transfected L-*myc* gene can substitute for c-*myc* in blocking murine erythroleukemia differentiation. Mol. Cell. Biol., *9:* 2734–2737, 1989.
16. Yokota, J., Wada, M., Yoshida, T., Noguchi, M., Terasaki, T., Shimosato, Y., Sugimura, T., and Terada, M. Heterogeneity of lung cancer cells with respect to the amplification and rearrangement of *myc* family oncogenes. Oncogene, *2:* 607–611, 1988.
17. Yokota, J., Wada, M., Shimosato, Y., Terada, M., and Sugimura, T. Loss of heterozygosity on chromosomes 3, 13, and 17 in small-cell carcinoma and on chromosome 3 in adenocarcinoma of the lung. Proc. Natl. Acad. Sci. USA, *84:* 9252–9256, 1987.
18. Miura, I., Siegfried, J. M., Resau, J., Keller, S. M., Zhou, J. Y., and Testa, J. R. Chromosome alterations in 21 non-small cell lung carcinomas. Genes Chromosomes Cancer, *2:* 328–338, 1990.
19. Whang-Peng, J., Knutsen, T., Gazdar, A., Steinberg, S. M., Oie, H., Linnoila, I., Mulshine, J. L., Nau, M., and Minna, J. D. Nonrandom structural and numerical chromosome changes in non-small cell lung cancer. Genes Chromosomes Cancer, *3:* 168–188, 1991.
20. Beardsley, T. Smart genes. Sci. Am., *265:* 87–95, 1991.
21. Schutte, J., Minna, J. D., and Birrer, M. J. Deregulated expression of human c-*jun* transforms primary rat embryo cells in cooperation with an activated c-Ha-*ras* gene and transforms rat-1a cells as a single gene. Proc. Natl. Acad. Sci. USA, *86:* 2257–2261, 1989.
22. Sunday, M. E., Hua, J., Dai, H. B., Nusrat, A., and Torday, J. S. Bombesin increases fetal lung growth and maturation *in utero* and in organ culture. Am. J. Respir. Cell. Mol. Biol., *3:* 199–205, 1990.
23. Kane, M. A., Aguayo, S. M., Portanova, L. B., Ross, S. E., Holley, M., Kelley, K., and Miller, Y. E. Isolation of the bombesin/gastrin-releasing peptide receptor from human small cell lung carcinoma NCI-H345 cells. J. Biol. Chem., *266:* 9486–9493, 1991.
24. Aguayo, S. M., Kane, M. A., King, T. E., Jr., Schwarz, M. I., Grauer, L., and Miller, Y. E. Increased levels of bombesin-like peptides in the lower respiratory tract of asymptomatic cigarette smokers. J. Clin. Invest., *84:* 1105–1113, 1989.
25. Cuttitta, F., Rosen, S., Gazdar, A. F., and Minna, J. D. Monoclonal antibodies that demonstrate specificity for several types of human lung cancer. Proc. Natl. Acad. Sci. USA, *78:* 4591–4595, 1981.
26. Mulshine, J. L., Cuttitta, F., Bibro, M., Fedorko, J., Fargion, S., Little, C., Carney, D. N., Gazdar, A. F., and Minna, J. D. Monoclonal antibodies that distinguish non-small cell from small-cell lung cancer. J. Immunol., *131:* 497–502, 1983.
27. Tockman, M. S., Gupta, P. K., Myers, J. D., Frost, J. K., Baylin, S. B., Gold, E. B., Chase, A. M., Wilkinson, P. H., and Mulshine, J. Sensitive and specific monoclonal antibody recognition of human lung cancer antigen on preserved sputum cells: a new approach to early lung cancer detection. J. Clin. Oncol., *6:* 1685–1693, 1988.
28. Hakomori, S. Biochemical basis and clinical application of tumor-associated carbohydrate antigens: current trends and future perspectives. Jpn. J. Cancer Chemother., *16:* 715–731, 1989.
29. Hakomori, S. Glycosphingolipids. Sci. Am., *54:* 44–53, 1986.
30. Fukushi, Y., Ohtani, H., and Orikasa, S. Expression of lacto series type 2 antigens in human renal cell carcinoma and its clinical significance. J. Natl. Cancer Inst., *81:* 352–358, 1989.

complex pattern of translational initiation and phosphorylation in L-myc proteins. Oncogene, *6:* 371–378, 1991.

32. Kyogashima, M., Mulshine, J., Linnoila, R., Jensen, S., Magnani, J. L., Nudelman, E., Hakomori, S., and Ginsburg, V. Antibody 624H12, which detects lung cancer at early stages, recognizes a sugar sequence in the glycosphingolipid difucosylneolactonorhexaosylceramide. Arch. Biochem. Biophys., *275:* 309–314, 1989.
33. Spitalnik, S. L., Spitalnik, P. F., Dubois, C., Mulshine, J., Magnani, J. L., Cuttitta, F., Civin, C. I., Minna, J. D., and Ginsburg, V. Glycolipid antigen expression in human lung cancer. Cancer Res., *46:* 4751–4755, 1986.
34. Gupta, P. K., Myers, J. D., Baylin, S. B., Mulshine, J. L., Cuttitta, F., and Gazdar, A. F. Improved antigen detection in ethanol-fixed cytologic specimens. A modified avidin-biotin-peroxidase complex (ABC) method. Diagn. Cytopathol., *1:* 133–136, 1985.
35. Huang, L. C., Brockhaus, M., Magnani, J. L., Cuttitta, F., Rosen, S., Minna, J. D., and Ginsburg, V. Many monoclonal antibodies with an apparent specificity for certain lung cancers are directed against a sugar sequence found in lacto-*N*-fucopentaose III. Arch. Biochem. Biophys., *220:* 318–320, 1983.
36. Combs, S. G., Marder, R. J., Minna, J. D., Mulshine, J. L., Polovina, M. R., and Rosen, S. T. Immunohistochemical localization of the immunodominant differentiation antigen lacto-*N*-fucopentaose III in normal adult and fetal tissues. J. Histochem. Cytochem., *32:* 982–988, 1984.
37. Hemminki, K., Perera, F. P., Phillips, D. H., Randerath, K., Reddy, V., and Santella, R. M. Aromatic deoxyribonucleic acid adducts in white blood cells of foundry and coke oven workers. Scand. J. Work Environ. Health, *14:* 55–56, 1988.
38. Lippman, S. M., Peters, E. J., Wargovich, M. J., Dixon, D. O., Dekmezian, R. H., Cunningham, J. E., Loewy, J. W., Morice, R. C., and Hong, W. K. The evaluation of micronuclei as an intermediate endpoint of bronchial carcinogenesis. Prog. Clin. Biol. Res., *339:* 165–177, 1990.
39. Stich, H. F. The use of micronuclei in tracing the genotoxic damage in the oral mucosa of tobacco users. *In:* H. Hoffman and C. Harris (eds.), Mechanisms in Tobacco Carcinogenesis, pp. 99–109. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
40. Liou, S. H., Jacobson Kram, D., Poirier, M. C., Nguyen, D., Strickland, P. T., and Tockman, M. S. Biological monitoring of fire fighters: sister chromatid exchange and polycyclic aromatic hydrocarbon-DNA adducts in peripheral blood cells. Cancer Res., *49:* 4929–4935, 1989.
41. Hulka, B. S., and Wilcosky, T. Biological markers in epidemiologic research. Arch. Environ. Health, *43:* 83–89, 1988.
42. Makuch, R. W., and Muenz, L. R. Evaluating the adequacy of tumor markers to discriminate among distinct populations. Semin. Oncol., *14:* 89–101, 1987.
43. Gail, M. H., Muenz, L., McIntire, K. R., Radovich, B., *et al.* Multiple markers for lung cancer diagnosis: validation of models for localized lung cancer. J. Natl. Cancer Inst., *80:* 97–101, 1988.
44. Kawajiri, K., Nakachi, K., Imai, K., Yoshii, A., Shinoda, N., and Watanabe, J. Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. FEBS Lett., *263:* 131–133, 1990.
45. Erozan, Y. S., and Frost, J. K. Cytopathologic diagnosis of lung cancer. *In:* M. J. Straus (ed.), Lung Cancer Clinical Diagnosis and Treatment, Ed. 2, pp. 113–125. New York: Grune & Stratton, 1983.
46. Eggleston, J. C., Tockman, M. S., Baker, R. R., Erozan, Y. S., Marsh, B. R., Ball, W. C., Jr., and Frost, J. K. *In-situ* and microinvasive squamous cell carcinoma of the lung. Clin. Oncol., *1:* 499–512, 1982.
47. Frost, J. K., Erozan, Y. S., Gupta, P. K., and Carter, D. Cytopathology. *In:* Atlas of Early Lung Cancer, sect. 3, pp. 39–70. New York and Tokyo: National Cancer Institute and Igaku-Shoin Medical Publishers, Inc., 1983.
48. Lippman, S. M., Lee, J. S., Lotan, R., Hittelman, W., Wargovich, M. J., and Hong, W. K. Biomarkers as intermediate end points in chemoprevention trials. J. Natl. Cancer Inst., *82:* 555–560, 1990.
49. Slaughter, D. P., Southwick, H. W., and Smejkal, W. "Field cancerization" in oral stratified squamous epithelium: clinical implications of multicentric origin. Cancer (Phila.), *6:* 963–968, 1953.
50. Fleiss, J. L. Statistical factors in early detection of health effects. *In:* D. W. Underhill and E. P. Radford (eds.), New and Sensitive Indicators of Health Impacts of Environmental Agents, pp. 9–16. Pittsburgh: University of Pittsburgh, 1986.
51. Peto, T. E., Thein, S. L., and Wainscoat, J. S. Statistical methodology in the analysis of relationships between DNA polymorphisms and disease: putative association of Ha-*ras*-1 hypervariable alleles and cancer. Am. J. Hum. Genet., *42:* 615–617, 1988.
52. Lilienfeld, A. M., and Lilienfeld, D. E. Foundations of Epidemiology, 2. Ed. pp. 133–165. New York: Oxford University Press, 1980.
53. Goyer, R. A., and Rogan, W. J. When is biologic change an indicator of disease? *In:* D. W. Underhill and E. P. Radford (eds.), New and Sensitive Indicators of Health Impacts of Environmental Agents, pp. 17–25. Pittsburgh: University of Pittsburgh, 1986.
54. Tockman, M. S. Development of labels of early lung cancer at the John K. Frost Center for Imaging of Cells and Molecular Markers. Lung Cancer Res. Q., *1:* 4–6, 1991.
55. Bartels, P. H., Bahr, G. F., Jeter, W. S., Olson, G. B., Taylor, J., and Wied, G. L. Evaluation of correlational information in digitized cell images. J. Histochem. Cytochem., *22:* 69–79, 1974.
56. Olson, G. B., Donovan, R. M., Bartels, P. H., Pressman, N. J., and Frost, J. K. Microphotometric differentiation of human T and B cells tagged with

monospecific immunoadsorbent beads. Anal. Quant. Cytol., 2: 144–152, 1980.

57. Frost, J. K., Pressman, N. J., Gill, G. W., *et al.* Quantitative nuclear analysis of developing squamous cell carcinoma of the human lung. Anal. Quant. Cytol., *5:* 207, 1983.

58. Tyrer, H. W., Frost, J. K., Pressman, N. J., *et al.* Automatic cell identification and enrichment in lung cancer: IV. Small cell carcinoma analysis by light scatter (size) and two fluorescence parameters (DNA, RNA). Flow Cytom., *4:* 464–472, 1980.

59. Pressman, N. J., Frost, J. K., Showers, R. L., *et al.* Measurement of spectral energy absorption in cytologic preparations using scanning transmission optical microscopy. Anal. Quant. Cytol., *4:* 155, 1982.

60. Bacus, J. W., Grace, L. J. Optical microscope system for standardized cell measurements and analyses. Appl. Optics, *26:* 3280–3293, 1987.

61. Oud, P. S., Henselaar, A. G. J. M., Pahlplatz, M. M. M., Meijer, J. W. R., and Vooijs, G. P. Image DNA-index (ploidy) analysis in cancer diagnosis. Appl. Optics, *26:* 3349–3355, 1987.

62. Wong, E. K., Liang, E. H., Lin, E. K., Simmons, D. A., and Koss, L. G. A selective mapping algorithm for computer analysis of voided urine cell images. Anal. Quant. Cytol. Histol., *11:* 203–210, 1989.

63. Kommoss, F., Bibbo, M., Colley, M., Dytch, H. E., Franklin, W. A., Holt, J. A., and Wied, G. L. Assessment of hormone receptors in breast carcinoma by immunocytochemistry and image analysis I. Progesterone receptors. Anal. Quant. Cytol. Histol., *11:* 298–306, 1989.

64. Cohen, O., Brugal, G., Seigneurin, D., and Demongeot, J. Image cytometry of estrogen receptors in breast carcinomas. Cytometry, *9:* 579–587, 1988.

65. Norazmi, M. N., Hohmann, A. W., Skinner, J. M., Jarvis, L. R., and Bradley, J. Density and phenotype of tumour-associated mononuclear cells in colonic carcinomas determined by computer assisted video image analysis. Immunology, *69:* 282–286, 1990.

66. Oud, P. S., Pahlplatz, M. M. M., Beck, J. L. M., Wiersma-van Tilburg Wagenaar, S. J., and Vooijs, G. P. Image and flow DNA cytometry of s cell carcinoma of the lung. Cancer (Phila.), *64:* 1304–1309, 1989.

67. Broers, J. L. V., Pahlplatz, M. M. M., Katzko, M. W., Oud, P. S., Ramael F. C. S., Carney, D. N., and Vooijs, G. P. Quantitative description of cl and variant small cell lung cancer cell lines by nuclear image cytom Cytometry, *9:* 426–431, 1988.

68. Weger, A. R., Mikuz, G., Askensten, U., Auer, G. U., Schwab, G., Glaser, K. S. Methodological aspects of DNA image cytometry in form fixed paraffin-embedded material from pancreatic adenocarcinoma.

69. Carruba, G., Pavone, C., Pavone-Macaluso, M., Mesiti, M., d'Aquino Vita, G., Sica, G., and Castagnetta, L. Morphometry of *in vitro* systems image analysis of two human prostate cancer cell lines (PC3 and DU- Pathol. Res. Pract., *185:* 704–708, 1989.

70. Gavoille, A., Kahn, E., Bosq, J., and Malki, M. B. A user-oriented soft for cytological image analysis application to automatic DNA content m urement of thyroid cells. Pathol. Res. Pract., *185:* 821–824, 1989.

71. Lesty, C., Raphael, M., Nonnenmacher, L., and Binet, J. L. Two stati approaches to nuclear shape and size in a morphometric description of ly node sections in non-Hodgkin's lymphoma. Cytometry, *10:* 28–36, 198

72. Pressman, N. J. Computer-aided microscopy in medicine: Introducti the 15 August 1987 issue of Applied Optics. Appl. Optics, *26:* 3199– 1987.

73. Greenberg, S. D., Spjut, H. J., Estrada, R. G., Hunter, N. R., and Gr C. Morphometric markers for the evaluation of preneoplastic lesions i lung. Diagnostic evaluation by high-resolution image analysis of aty cells in sputum specimens. Anal. Quant. Cytol. Histol., *9:* 49–54, 1987

74. Thomas, P. A., and Piantadosi, S. Postoperative T1 N0 non-small cell cancer. J. Thorac. Cardiovasc. Surg., *94:* 349–354, 1987.